

Universidad de San Andrés
Biblioteca Max von Buch
Colecciones Especiales y Archivos

Estándares de digitalización y preservación de imágenes de la Colección Hume Hermanos

Silvana Lucía Piga
Natalia Westberg

Septiembre 2014

Agradecimientos

Por la colaboración de María Esteva, Santiago Mancardo, Francisco Moreno y Lucila Quercia.

Estándares de digitalización y preservación de imágenes de la Colección Hume Hermanos

Este documento de trabajo fue elaborado en relación al proyecto de organización, digitalización y preservación de la colección **Hume Hermanos**. El mismo fue financiado por la **British Library** a través del [Endangered Archives Programme](#). Se vio la necesidad de implementar un programa de preservación digital para que el material digitalizado fuera accesible a largo plazo. Para esto último se contó con el apoyo del **Ministerio de Ciencia, Tecnología e Innovación Productiva de la Nación** a través de la beca **César Milstein** del [Programa Raíces](#), que permitió la visita de la **Dra. Maria Esteva**, quien capacitó al personal y guió la implementación de un ciclo de vida digital. Este ciclo de vida permite la organización de los objetos digitales desde su origen en la digitalización, automatiza tareas de documentación técnica y descriptiva de la colección digital, y asegura la carga de objetos digitales auténticos a un repositorio **DSpace**.

Manual de Procedimientos

Aquí se detallan las tareas relacionadas con la planificación del proyecto, se describen los procedimientos aplicados y se presenta el flujo de trabajo semi-automatizado que fueron implementados para hacer más precisa y eficiente la tarea. Se incorpora también bibliografía y enlaces que fueron de mucha utilidad durante el proyecto.

El manual está organizado en los siguientes apartados:

I. Planificación

1.-Análisis de la colección

2.-Organización de la colección

2.1.- Nombramiento de archivos digitales y de la estructura del directorio de guarda

3.-Mapeo de metadatos

4.-Correspondencia de la colección con la arquitectura de DSpace

II. Procedimientos

5.-Limpieza, tratamiento y guarda de los documentos originales/analógicos

6.-Digitalización de los documentos originales

6.1.- Tipos de archivos generados

6.2.- Parámetros de la digitalización

6.3.- Perfilado del escáner

6.4.- Calibración del monitor

6.5.- Escaneado

6.6.- Control de calidad

7.-Creación de ítems (grupos de archivos) para importación en Dspace

7.1.- Generación de archivos de preservación

7.1.1.- Generación del master de archivo

7.1.2.- Generación de metadatos técnicos

7.2.-Generación de metadatos descriptivos

7.3.-Generación de archivos de publicación

7.4.-Generación de archivo para licencia

7.5.-Carga del bundle

8.-Flujo final de trabajo

9.-Medidas de preservación: réplica o back up

10.-Softwares y estándares utilizados en el proyecto

11.-Links relacionados con el proyecto institucional de digitalización

12.-Bibliografía

I.- Planificación

1.- Análisis de la colección

El objetivo de este paso es comprender la estructura y el contenido de la colección estableciendo su abarque temporal, su organización y su tipología documental. Para ello se realizaron las siguientes tareas:

- Determinar la estructura organizativa y el orden original de la colección Hume Hermanos
- Identificar tipos documentales (ej. álbumes, placas de vidrio, cuadernos, fotografías sueltas, etc.)
- Identificar grupos de documentos y su temática común
- Identificar puntos descriptivos para el acceso a la colección
- Inventariar el material
- Releva su estado de conservación
- Planificar las intervenciones de conservación requeridas antes de la digitalización (ej.: alisado, limpieza, encuadernados)
- Verificar el tamaño de los documentos para que ocupen cómodamente la superficie de la platina del escáner
- Determinar la cantidad de documentos a digitalizar (ej. cantidad de fotografías, placas de vidrio, cuadernos, álbumes)

2.- Organización de la colección

A partir del análisis se determinó que la colección Hume Hermanos presenta la siguiente estructura:

Serie 1: Documentos legales y correspondencia

Serie 2: Documentos profesionales

Serie 3: Documentos técnicos y proyectos

Serie 4: Copiadores de correspondencia comercial

Serie 5: Fotografías

Serie 6: Memorias de la empresa y la familia

En este proyecto, la organización de la colección digital está dada por el orden original de la colección física. Para el almacenamiento de los archivos digitales, la organización de la colección se trasladó a un directorio jerárquico virtual. (Figura 1). Se denomina ítem a cada unidad documental compuesta por una o más imágenes como, por ejemplo, una fotografía suelta, un libro copiadore o un álbum fotográfico. Las series son, en esta colección, grupos de ítems relacionados.

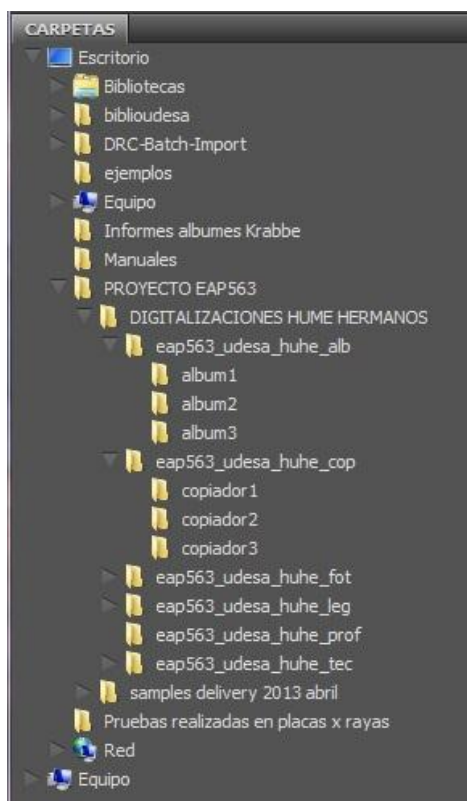


Figura 1: Estructura de la colección representada en un directorio virtual.

2.1.- Nombramiento de archivos digitales y de la estructura del directorio de guarda

El nombramiento de la estructura del directorio y de las imágenes es clave para extraer metadatos útiles programáticamente. Los directorios se etiquetaron con un sistema de código que incluye numeración y términos para describir la procedencia y los contenidos de la colección.

Para el nombramiento de las imágenes que se encuentran dentro de los directorios se utilizó el siguiente criterio:

- Cada nombre debe ser único, aunque la mayor parte del nombre conserva elementos comunes para grupos de imágenes que conforman un ítem (ej. todas las imágenes correspondientes a un copiador)
- Cada nombre debe tener un elemento que lo haga único (ej. un número único secuencial)
- Debe mantenerse la uniformidad a lo largo de los diferentes ítems de la colección
- El esquema de nombramiento debe tener una secuencia lógica que pueda ser fácilmente identificada
- No utilizar nombramientos complejos o muy largos que sean susceptibles de generar errores humanos durante el tipeado
- Seguir las mejores prácticas para la selección de caracteres al crear los nombramientos:
 - No usar espacios en blanco

- Usar guiones bajos o medios
- Evitar otro tipo de símbolos

El esquema de nombramiento de archivos se terminó de definir contemplando los elementos requeridos por la **British Library**. Por ejemplo, la British Library exigía que figurara la identificación del proyecto (eap563) y la pertenencia institucional (udesa). Luego de definir el sistema de nombramiento se confeccionó un diccionario que explica los códigos y abreviaturas como puede verse en el siguiente ejemplo:

Nombre de un archivo digital: eap563_udesa_huhe_cop1_006.tiff

Nombre del proyecto	Sigla de la institución	Abreviatura de la colección	Serie: copiadore y subserie : copiador N° 1	numeración secuencial automática	Formato
eap563	udesa	huhe	cop1	006	tiff
eap563	udesa	huhe	cop1	<u>007</u>	tiff
eap563	udesa	huhe	cop1	<u>008</u>	tiff
eap563	udesa	huhe	cop1	<u>009</u>	tiff

Para nombrar los archivos de imágenes se utilizó el software **Silver Fast**. La función del **Silver Fast** es tomar el esquema de nombramiento, aplicarlo a un grupo de imágenes que forman un ítem, y asignar automáticamente un número correlativo a cada imagen como puede verse en la figura 2. Este número identifica cada imagen de modo único y el esquema de nombramiento relaciona a todas las imágenes que conforman un ítem, así como a todos los ítems que forman una serie.

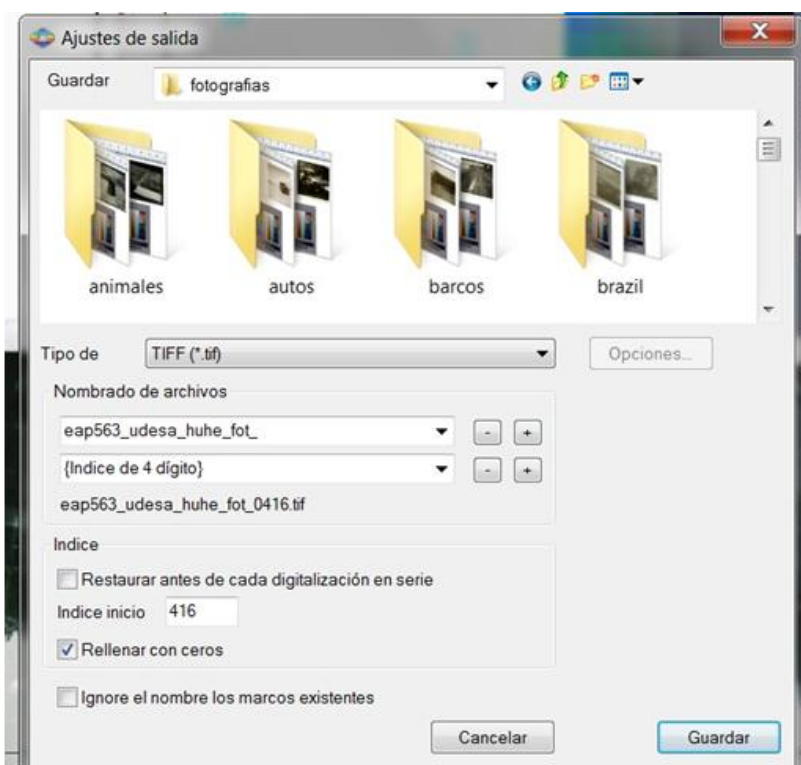


Figura 2: Asignación del nombre del archivo digital en Silver Fast

3.-Mapeo de metadatos

Con el objetivo de automatizar el proceso de generación de metadatos y exportarlos al repositorio *DSpace*, primero se realizó un mapeo de la información presente en el esquema de nombramiento del directorio y de las imágenes hacia los elementos de un estándar de metadatos descriptivos. El mapeo permite traducir los elementos del nombre y del directorio entre los distintos lenguajes y establecer la correspondencia entre uno y otro (Figura 3). En este proyecto se decidió usar **Dublin Core (DC)** porque se trata de un estándar sencillo y, porque es soportado por la mayoría de los software libres de repositorios digitales, como por ejemplo *DSpace*. El **DC** se emplea para codificar metadatos de cada ítem de la colección en un documento **XML** (Figura 18). Luego, el documento **DC/XML** se adjunta con las imágenes que corresponden al ítem y son utilizados por el software *DSpace* para poblar automáticamente los campos descriptivos en el repositorio. En la Figura 3 se muestra el mapeo de elementos extraídos del nombre del directorio y del archivo a **DC** y a **EAD (Encoded Archival Description)**. Por ejemplo, se ve que en el elemento *sujeto* de **DC**, se puebla con el término “fotografías,” término de la etiqueta descriptiva de la Serie V.

Item: Copiador 1				
Elemento EAD	Elemento Dublin Core	Ejemplo	Extraído de	
<subject>	<subject>	<dcvalue qualifier="none" element="subject"> copiadores </dcvalue>	nombre del directorio	C:\Users\biblioudesa\Desktop\Proyecto eap563\hume hermanos\copiadores_correspondencia_comercial\copiador1
<subject>	<subject>	<dcvalue qualifier="none" element="subject"> correspondencia </dcvalue>	nombre del directorio	C:\Users\biblioudesa\Desktop\Proyecto eap563\hume hermanos\copiadores_correspondencia_comercial\copiador1
<subject>	<subject>	<dcvalue qualifier="none" element="subject"> comercial </dcvalue>	nombre del directorio	C:\Users\biblioudesa\Desktop\Proyecto eap563\hume hermanos\copiadores_correspondencia_comercial\copiador1
<title>	<title>	<dcvalue qualifier="none" element="title"> copiador1 </dcvalue>	nombre del directorio	C:\Users\biblioudesa\Desktop\Proyecto eap563\hume hermanos\copiadores_correspondencia_comercial\copiador1
<unitid>	<identifier:other>	<dcvalue qualifier="other" element="identifier"> eap563_udesa_huhe_cop1 </dcvalue>	nombre del archivo	eap563_udesa_huhe_cop1
<publisher>	<publisher:digital>	<dcvalue qualifier="digital" element="publisher"> udesa </dcvalue>	nombre del archivo	eap563_udesa_huhe_cop1
<sponsor>	<description:sponsorship>	<dcvalue qualifier="sponsorship" element="description"> eap563 </dcvalue>	nombre del archivo	eap563_udesa_huhe_cop1
<c level="collection">	<relation:ispartof>	<dcvalue qualifier="ispartof" element="relation"> hume hermanos </dcvalue>	nombre del directorio	C:\Users\biblioudesa\Desktop\Proyecto eap563\hume hermanos\copiadores_correspondencia_comercial\copiador1
<c level="series">	<relation:ispartof>	<dcvalue qualifier="ispartof" element="relation"> copiadores_correspondencia_comercial </dcvalue>	nombre del directorio	C:\Users\biblioudesa\Desktop\Proyecto eap563\hume hermanos\copiadores_correspondencia_comercial\copiador1
<c level="item">	<relation:haspart>	<dcvalue qualifier="haspart" element="relation"> eap563_udesa_huhe_cop1_0001 </dcvalue>	nombre del archivo	eap563_udesa_huhe_cop1_0001
<author>	<creator>	<dcvalue qualifier="none" element="creator"> huhe </dcvalue>	nombre del archivo	eap563_udesa_huhe_cop1
<controlaccess><genreform>	<type>	<dcvalue qualifier="none" element="type"> Images </dcvalue>	extraído con fits	
	<format>	<dcvalue qualifier="mimetype" element="format"> tiff </dcvalue>	extraído con fits	
	<relation:hasformat>	<dcvalue qualifier="hasformat" element="relation"> eap563_udesa_cop1_001.jpg </dcvalue>	nombre del archivo	

Figura 3: Mapeo de metadatos

4.-Correspondencia de la colección con la arquitectura de DSpace

La arquitectura del repositorio DSpace debe reflejar la estructura de la colección. En **Dspace**, los objetos digitales se organizan en comunidades, sub-comunidades, colecciones, e ítems. La figura 4 muestra la correspondencia entre las series e ítems de la colección Hume y la arquitectura de **Dspace**.

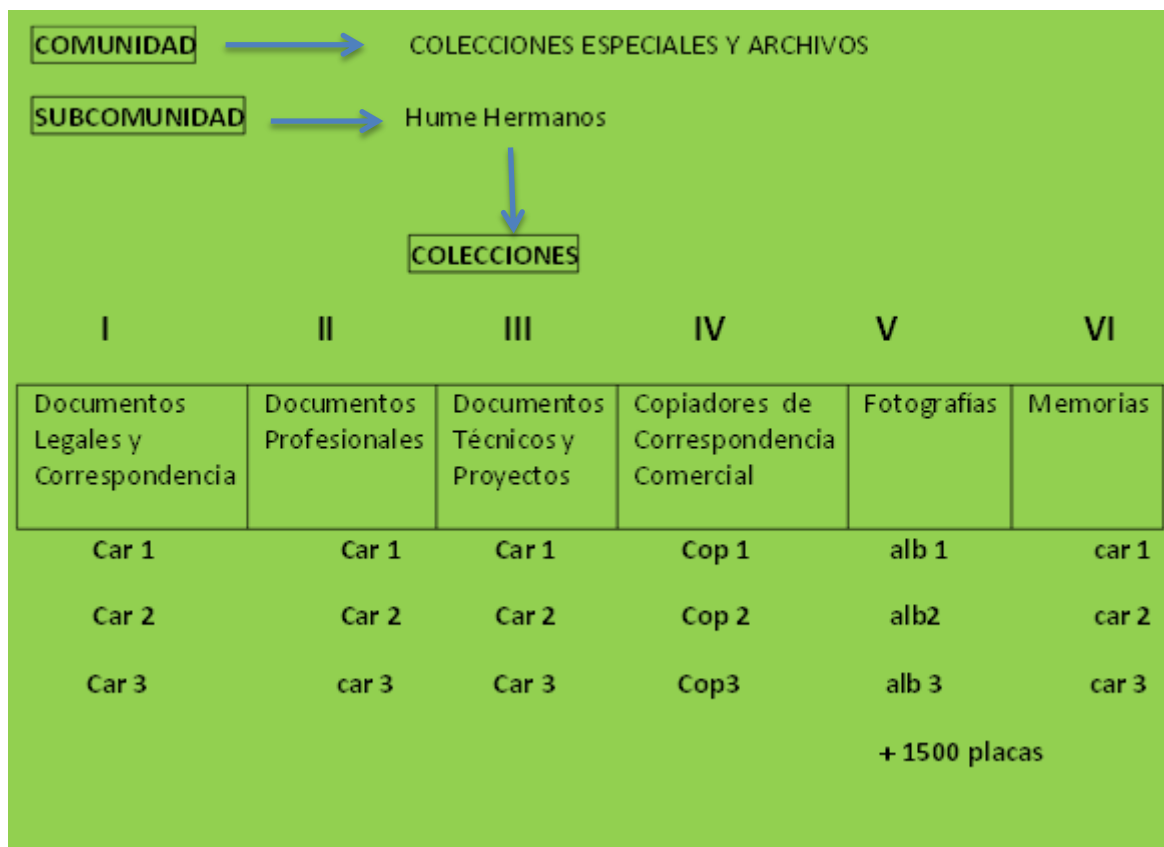


Figura 4: Organización de la Colección Hume Hermanos en el repositorio DSpace

La Biblioteca tiene algunas colecciones en el repositorio digital de **Dspace**. Cada colección recibe el nombre de Comunidad. El sector Colecciones Especiales y Archivos creó su propia comunidad dentro de la cual cada uno de los fondos documentales será una subcomunidad. A su vez las series que la componen recibirán el nombre de colecciones. Cada colección (serie) puede contener múltiples ítems.

II. Procedimientos

5.- Limpieza, tratamiento y guarda de los documentos originales/analógicos:

El primer paso comprende estabilizar la colección original, prepararla para digitalizarla y para su guarda de preservación. La colección Hume Hermanos ingresó al sector Colecciones Especiales y Archivos con algunos materiales en muy mal estado de conservación. Entre ellos se destacan las placas de vidrio estereoscópicas. Para realizar los tratamientos de limpieza es necesario tener formación en conservación de papel y en el caso específico de las fotografías, contar con entrenamiento en tareas básicas de limpieza de material fotográfico. Tareas más avanzadas como la consolidación de la emulsión y tratamiento de hongos en las placas de vidrio, fueron realizadas por una conservadora profesional de fotografías. Los contenedores para las placas fueron realizados a medida. A continuación se enuncian las tareas necesarias para estabilización de la colección:

Papel:

- Limpieza mecánica con goma de borrar y pinceleta.
- Extracción de ganchos metálicos y alfileres.
- Alisado de pliegues y dobleces.
- Guarda en contenedores libres de ácido.



Figura 5: Documentos limpios almacenados en carpetas y cajas libres de ácido

Fotografías:

- Limpieza de placas de vidrio y de fotografías en papel.
- Consolidación de las emulsiones dañadas.
- Guarda en contenedores libres de ácido.

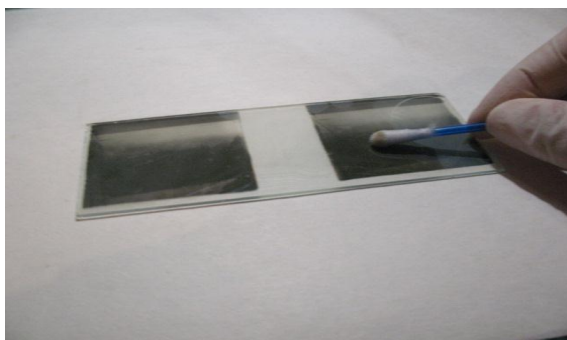
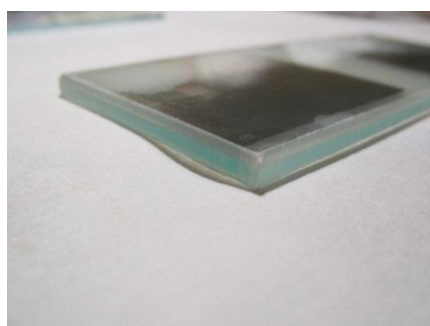
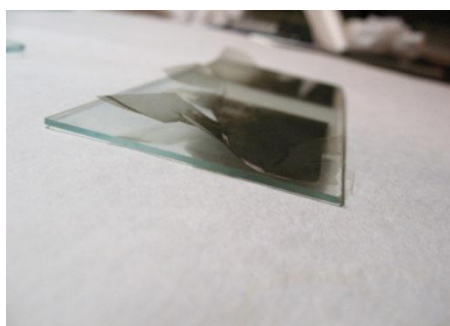


Figura 6: Tratamiento de limpieza de una placa de vidrio estereoscópica



Figuras 7 y 8: Placa con emulsión desprendida y placa tratada



Figura 9: Placas de vidrio ensobradas con papel neutro dentro de contenedor de polipropileno confeccionado a medida

6.- Digitalización de los documentos originales

6.1.- Tipos de archivos generados

En los proyectos de digitalización pueden generarse tres tipos de archivos digitales: master de archivo, master de publicación y derivados de acceso.

Se describe brevemente a continuación la función de cada uno de ellos y explica el criterio adoptado para la digitalización.

Master de archivo: el archivo generado directamente desde el escáner, sin compresión, de alta resolución en formato **TIFF**.

Master de publicación: creado a partir del master de archivo que es retocado para asemejar el tono y el color a la imagen física. Se guarda con el retoque fotográfico en formato **TIFF**.

Derivado de acceso: archivo comprimido, de menor resolución y tamaño que se utiliza para presentación web. De este modo el archivo puede descargarse con facilidad teniendo en cuenta las limitaciones de banda.

Se decidió no generar un master de publicación. Los retoques hechos en las imágenes son guardados directamente como derivados de acceso. El motivo de esta decisión se relaciona con una cuestión de falta de espacio en el servidor. A partir del archivo **TIFF** se crea como derivados imágenes **JPG**. La conversión de **TIFF** a **JPG** se realiza con el programa **Photoshop Elements**. En el caso de que un usuario requiriera la imagen de archivo **TIFF** se la haría llegar en una forma conveniente.

6.2.- Parámetros de la digitalización

Para definir estos parámetros se tomaron en consideración las directrices de la **British Library** para la digitalización de cada tipo de material estableciéndose el siguiente cuadro:

Imagen Digitalizada	Texto impreso	Texto manuscrito (MS)	Fotografía en papel	Placa de vidrio
Original	Opaco	Opaco	Opaco	Transparencia
Pos/Neg	Positivo	Positivo	Positivo	Positivo
Formato	TIFF	TIFF	TIFF	TIFF
Tipo de Imagen	24-bits RGB	24-bits RGB	24- bits RGB	24- bit RGB
Resolución	400	300	400-600	1100

Figura 10: Parámetros seleccionados para la digitalización de los materiales

Diversos tamaños de originales requieren diversas resoluciones. La resolución se mide en **dpi** (*dots per inch*). Los originales más grandes pueden tener menor cantidad de dpi sin por ello pixelarse, perder nitidez o calidad al ampliarse. No es lo mismo digitalizar por ejemplo, un original de 10cm por 10 cm (una fotografía pequeña) que uno de 30cm x 30 cm. Para saber qué resolución requerirá cada tamaño de original se utilizó la fórmula que aparece en la Figura 11.

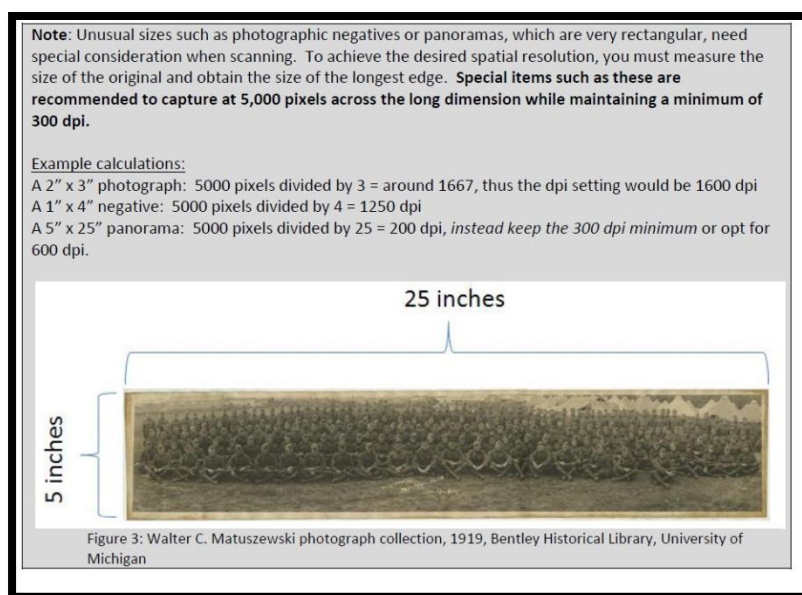


Figura 11: Ejemplo del cálculo para determinar la resolución. (Fuente: Digitization Guidelines for Photographs and Textual Documents. Julianna Barrera-Gomez. University of Michigan, July 2012)

6.3 Perfilado del escáner

Para realizar el trabajo de digitalización se eligió el escáner **EPSON 10000 XL** por considerar que se adaptaba mejor a las necesidades de la colección (imágenes reflectantes y transparencias, en este caso). Además se conocían buenas experiencias en el país con ese equipo.

Es necesario perfilar el escáner, es decir, programarlo estableciendo previamente los parámetros de ajustes de color y compensación de luminosidad que necesita para trabajar. Este perfilado se hace mediante el uso del software **Silver Fast**. El manual del software va guiando la selección de los parámetros más adecuados para el equipo y el material con el que se trabaja. En esta instancia se realizaron algunas consultas con expertos en el área de digitalización para respetar parámetros profesionales.

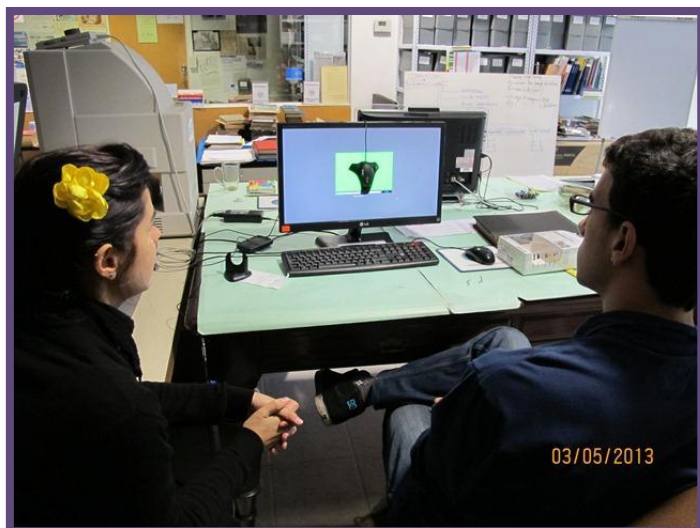


Figura 12: Proceso de calibración del monitor

6.4.- Calibración del monitor

La calibración del monitor garantiza un control de calidad homogéneo para todas las imágenes. Se utilizó **Spyder 4 PRO 4.5.4 (ver Anexo)**. El monitor de la computadora de trabajo debe ser calibrado periódicamente. Durante este proceso se evitó trabajar con luz blanca sobre el monitor y se mantuvieron las cortinas cerradas. Es importante chequear la limpieza de la pantalla y asegurarse que el monitor sea la fuente de luz más brillante en la sala.

6.5.- Escaneado

Se trabajó por lote y por ítem (por ejemplo, todas las imágenes correspondientes a un álbum de fotografías) con el fin de ahorrar tiempo, evitar la reprogramación del escáner cada vez y de permitir la automatización posterior de los pasos de extracción de metadatos e importación a **DSpace**.

Uno de los requisitos de la **BL** es el uso de regla y color checker en cada una de las imágenes. La regla sirve para que quienes consulten los archivos puedan conocer el tamaño del original. El color checker es una carta de colores que permite un control visual del proceso de digitalización.

6.6.- Control de calidad de las imágenes

Se trata de la inspección y evaluación sistemática y consistente de las imágenes digitales al lado de los objetos físicos/originales y de acuerdo a parámetros de calidad establecidos *a priori*, con el objetivo de garantizar una producción de imágenes digitales estable y uniforme. El control de calidad incluye algunos pasos que son simplemente la constatación de ciertos valores objetivos, como dimensiones y peso del archivo y otros que se sostienen en una observación subjetiva de ciertos elementos en la imagen. A pesar de depender del ojo humano, puede garantizarse un

resultado de calidad constante en la medida que el equipo de trabajo establezca los criterios y pautas que dirijan tal observación.

Durante el control se recomienda:

- Usar una luz focal para mirar los originales
- Mirar los originales sobre un fondo gris
- Usar vestimenta de colores neutros como gris, negra o blanca
- Visualizar la imagen digital en el monitor al 200%
- Evaluar una muestra (en porcentaje) de las imágenes producidas. Si hubiera detección de imperfecciones en la mayoría de las imágenes muestreadas se inspeccionará el 100%. En este proyecto se realizó una evaluación del 10% de los textos pero por ser una cantidad menor, se controló el 100% de las fotografías.

Verificar que:

- La ruta del archivo sea la correcta, por ejemplo:
C:\Users\biblioudesa\Desktop\Proyecto eap563\hume hermanos\fotografias\trenes
- Estén presentes y correctos todos los elementos del nombre del archivo
- Las imágenes estén en formato **TIFF**
- La resolución este de acuerdo a los parámetros establecidos para cada tipo de material
- El tamaño sea el correcto (de acuerdo con el establecido por la **BL**)
- El peso sea entre 30 y 40MB (de acuerdo con lo establecido por la **BL**)
- La imagen no esté invertida o rotada y esté bien centrada
- El color checker y la regla se vean completos, sin tapar el objeto digitalizado (Figura 16)
- El color y el tono sean fieles al original
- El contraste y el brillo sean similares a los del original
- No se encuentre pixelado ni efecto Moiré (Figuras 13 y 14)
- No se visualicen los anillos de Newton (Figura 15)



Figura 13: Pixelado

Fuente: www.fotosigno.com



Figura 14: Efecto Moiré

Fuente: www.conloschosentidos.com

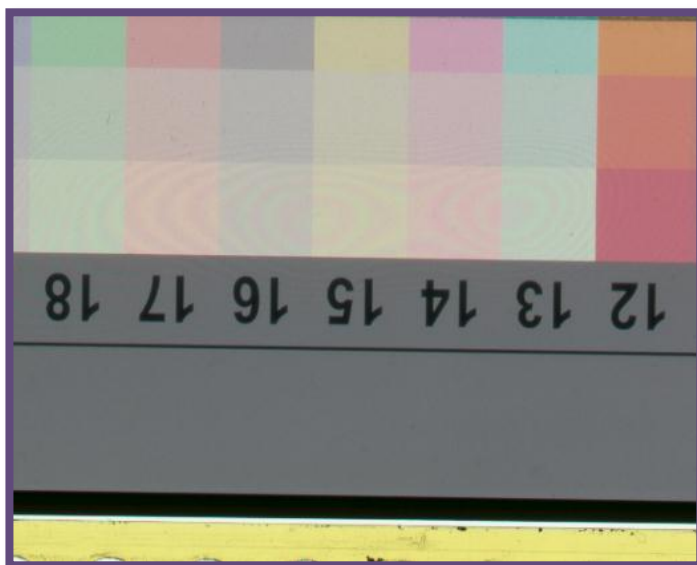


Figura 15: Anillos de Newton

Fuente: <http://www.jpereira.net/apuntes-breves/anillos-de-newton-en-escaneres>

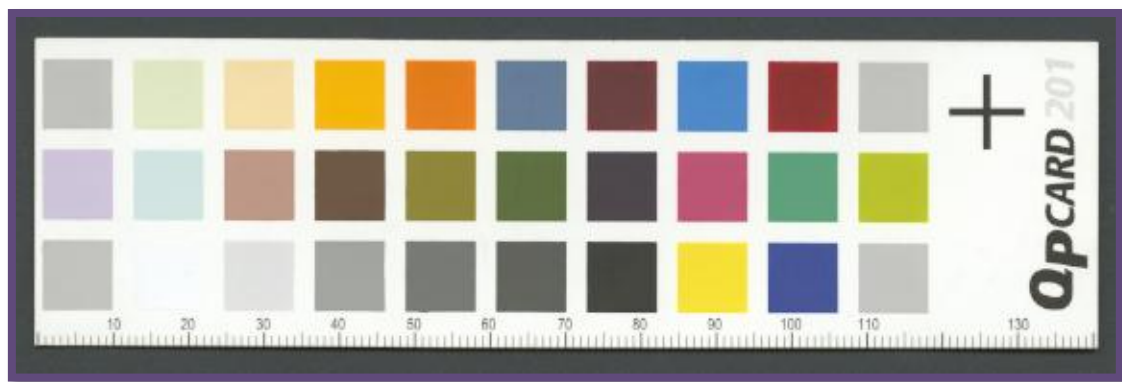


Figura 16: Color Checker y regla

Para documentos de texto evaluar en especial los siguientes aspectos:

- Que sea legible hasta en los caracteres más pequeños
- Que esté completo
- Contraste y nitidez de texto sobre fondo
- Alineación de los textos con respecto a los márgenes y uniformidad entre las líneas del texto

7.- Creación de ítems (grupos de archivos) para importación a Dspace

Dado que se seleccionó el software libre **Dspace** como repositorio para esta colección, fue necesario adaptar algunos criterios de la organización de la colección a la arquitectura del repositorio. En **Dspace** los archivos se almacenan como ítems, y cada ítem tiene una descripción en formato **DC**. Como se mencionó, los ítems son unidades de información y están compuestos por todas las imágenes/archivos que componen esa unidad de información. Es decir que un ítem de la colección puede ser una fotografía, un documento, un libro copiador o un álbum. Cada uno de estos ítems cuenta con su descripción particular. También se describieron aspectos generales del archivo y de cada una de las series en relación a sub-comunidades y colecciones. Estas descripciones se corresponden a descripciones de alcance y contenido, y series en **ISAD (G)**

7.1 Generación de archivos de preservación

Este proceso se compone de dos partes:

7.1.1 Generación del máster de archivo

Se trata del archivo **TIFF** generado por el escáner tal como se vio en el punto 6.1.

7.1.2 Generación de metadatos técnicos

Los metadatos técnicos son extraídos por un software llamado **FITS** (Flexible Image Transport System) y recopilados en un archivo **FITS.XML**. Este archivo es almacenado con cada objeto digital e indica resolución, tamaño, color, etc., siendo esta información imprescindible tanto para determinar la autenticidad del documento como para asegurar la estabilidad de la información en futuros procesos de migración, es decir, es un paso clave en la preservación digital.

7.2 Generación de metadatos descriptivos

Para la importación de cada ítem al sistema **DSpace** es necesario generar un archivo con sus metadatos descriptivos utilizando el estándar **DC**. Para generar este archivo para cada ítem en forma masiva se utilizó un programa creado por el informático que trabaja en la Biblioteca de nuestra Universidad. Este programa utiliza a su vez otras aplicaciones y automatiza los distintos pasos de un flujo de trabajo que va desde que las imágenes terminaron de ser digitalizadas y son ubicadas en un directorio jerárquico, pasando por la creación de metadatos y hasta el depósito de los ítems con su documentación completa en **Dspace**.

Este programa extrae los metadatos de la estructura de directorio y nombramiento de archivos, utilizando como modelo un mapeo a **DC** (Figura 3), puebla una aplicación macro en Excel elaborada por la Universidad de Ohio. El documento **DC** en xml luego se importa a **DSpace** (ver punto 7.5).

7.3 Generación de archivos de publicación y/o derivados de acceso

Se trata de los archivos **JPG**, creados a partir de los archivos **TIFF**, mencionados también en el punto Esto fue explicado en el punto 6. 1

7.4 Generación de archivos para licencia

También es necesario, para la publicación, un archivo con la licencia de uso del objeto digital. Se trata de un archivo en formato.txt. En el caso de este proyecto se decidió usar una licencia estándar de **Creative Commons**, denominada Reconocimiento-NoComercial-SinObraDerivada.

7.5.- Carga del bundle

Todos los archivos de metadatos generados en los procesos mencionados componen, junto con las imágenes que conforman un ítem, lo que se denomina *bundle*, el cual será ingresado a **DSpace**. Bundle significa manojo o ramillete y es el conjunto de archivos que componen el ítem. En este *bundle* se encuentran tanto los archivos de preservación como los de publicación (Figura 17) La importación de los *bundles* a **DSpace** se realiza en varios pasos. Para hacer este proceso más eficiente, es posible importar más de un *bundle* por vez.

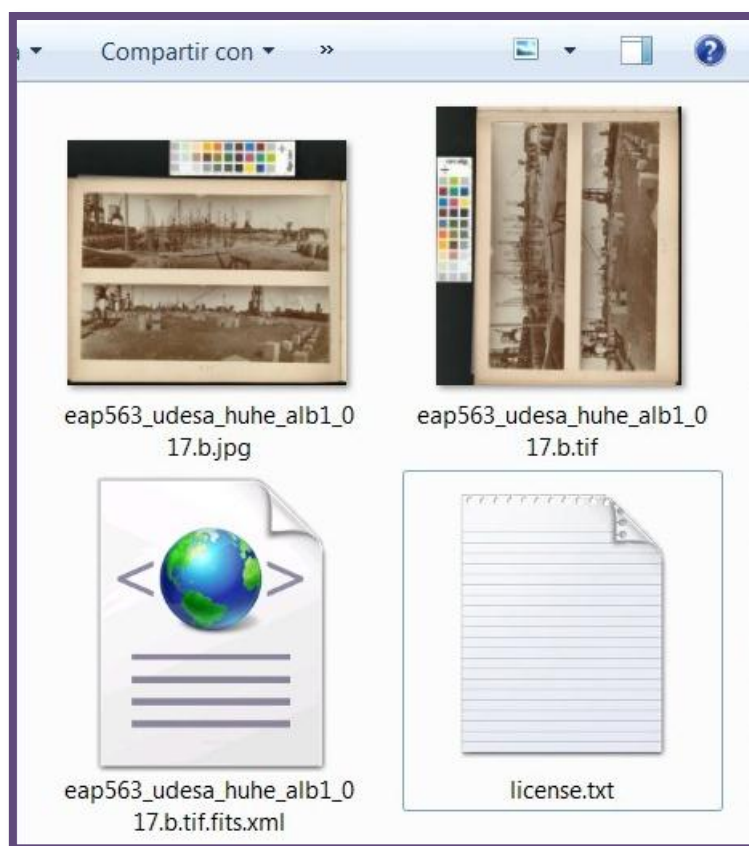


Figura 17: Estructura del bundle de un objeto de la colección

El primer paso es copiar la estructura de directorios que forman los *bundles* a un lugar temporal para acomodarla a la manera que la necesita **DSpace**. Esto se hace mediante un programa creado por el informático.

En el segundo paso el programa crea el *bundle* correspondiente a cada ítem. Este *bundle* está compuesto por:

- archivos de preservación: son el **TIFF** (master de archivo) y el **FITS.XML** (archivo de metadatos técnicos).
- archivos de publicación: El archivo **JPG** y el archivo de texto que contiene la licencia
- content.txt: archivo que especifica qué tipo de acceso tendrá cada archivo digital
- dublin_core.xml: estándar utilizado por **DSpace** para contener los metadatos descriptivos

Los archivos content.txt y dublin_core.xml (Figuras 18 y 19) son generados a partir de los datos generados en el punto 7.1 y utilizando una macro del archivo Excel creado y publicado por [Ohio LINK Digital Resources Management Committee](#). Este macro permite la extracción automática de los metadatos. La forma de utilizarlo se explica en el mismo archivo Excel de Ohio, en idioma inglés.

```

<?xml version="1.0" encoding="UTF-8" ?>
<: <dublin_core>
<dcvalue element="subject" qualifier="none">albumes</dcvalue>
<dcvalue element="subject" qualifier="none">fotografias</dcvalue>
<dcvalue element="title" qualifier="none">Album3</dcvalue>
<dcvalue element="identifier" qualifier="other">eap563_udesa_huhe_alb3</dcvalue>
<dcvalue element="publisher" qualifier="digital">udesa</dcvalue>
<dcvalue element="description" qualifier="sponsorship">eap563</dcvalue>
<dcvalue element="relation" qualifier="ispartof">huhe hermanos</dcvalue>
<dcvalue element="relation" qualifier="ispartof">albumes fotografias</dcvalue>
<dcvalue element="relation" qualifier="haspart">eap563_udesa_huhe_alb3_001.a</dcvalue>
<dcvalue element="relation" qualifier="haspart">eap563_udesa_huhe_alb3_001.b</dcvalue>
<dcvalue element="relation" qualifier="haspart">eap563_udesa_huhe_alb3_002.a</dcvalue>
<dcvalue element="relation" qualifier="haspart">eap563_udesa_huhe_alb3_002.b</dcvalue>
<dcvalue element="relation" qualifier="haspart">eap563_udesa_huhe_alb3_003.a</dcvalue>
<dcvalue element="relation" qualifier="haspart">eap563_udesa_huhe_alb3_003.b</dcvalue>
<dcvalue element="relation" qualifier="haspart">eap563_udesa_huhe_alb3_004.a</dcvalue>
<dcvalue element="relation" qualifier="haspart">eap563_udesa_huhe_alb3_004.b</dcvalue>
<dcvalue element="creator" qualifier="none">huhe hermanos</dcvalue>
<dcvalue element="type" qualifier="none">Images</dcvalue>
<dcvalue element="format" qualifier="mimetype">tiff</dcvalue>
</dublin_core>

```

Figura 18: Archivo XML/Dublin Core

Los usuarios estarán habilitados para visualizar los JPG y el archivo xml con los metadatos descriptivos. Los metadatos técnicos y el TIFF no estarán disponibles para los usuarios en la plataforma **DSpace**

En el siguiente paso, con el *bundle* completo, se ejecuta el programa de importación de **DSpace** indicando a qué colecciones pertenece. Puede exportarse más de un *bundle* cada vez.

Durante la importación a **DSpace** se realiza un Checksum de cada uno de los archivos. El checksum es una suma de verificación, que permite detectar cambios en una secuencia de datos. Esto hace posible en el futuro verificar la autenticidad y la integridad de los objetos.

Luego de esta etapa es importante el control de calidad de los metadatos que quedan registrados en el repositorio. Hay que revisar que todos los elementos contengan los valores correctos. Estos metadatos pueden ser corregidos y ampliados directamente desde la interfaz de **DSpace**. Figura 19

Webpage Screenshot



Repositorio Digital San Andrés

Buscar en el repositorio

Búsqueda avanzada

Registrado como

fuoreno@udesa.edu...

Salir

Página de inicio

Listar

Fecha Publicación

Autor

Título

Materia

Comunidades

Servicios

Alertas

Mi repositorio

usuarios autorizados

Editar perfil

Administrar

Ayuda

Sobre DSpace

creative commons

SOME RIGHTS RESERVED

Por favor, use este identificador para citar o enlazar este ítem: <http://hdl.handle.net/10908/847>

Exportar ítem

Export (migrate) ítem

Editar...

Export metadata

Registro completo de metadatos

Campo DC	Valor	Lengua/dioma
dc.creator hume hermanos		es
dc.date.accessioned 2013-05-14T13:16:06Z		-
dc.date.available 2013-05-14T13:16:06Z		-
dc.date.issued 2013-05-14		-
dc.identifier.other eap563_udesa_huhe_alb3		es
dc.identifier.uri http://hdl.handle.net/10908/847		-
dc.description.abstract esta es una pagina del album		-
dc.description.provenance Made available in DSpace on 2013-05-14T13:16:06Z (GMT). No. of bitstreams: 25 eap563_udesa_huhe_alb3_001.a.jpg: 3054290 bytes, checksum: f269228835c268a1138e7e058c2a2d59 (MD5) eap563_udesa_huhe_alb3_001.b.jpg: 2339136 bytes, checksum: 562319b9881ef20b22fc12ff4f15787 (MD5) eap563_udesa_huhe_alb3_002.a.jpg: 2036220 bytes, checksum: a75ac71fd03498195f3575cae181134a (MD5) eap563_udesa_huhe_alb3_002.b.jpg: 1984004 bytes, checksum: d1808e57832ec06c304a84997ce6dd88 (MD5) eap563_udesa_huhe_alb3_003.a.jpg: 2230434 bytes, checksum: 63d355a669ac9ff0fe98db0d30c4ec75 (MD5) eap563_udesa_huhe_alb3_003.b.jpg: 2008740 bytes, checksum: e166de02f5f8fe14371093d89f64c499 (MD5) eap563_udesa_huhe_alb3_004.a.jpg: 2051188 bytes, checksum: f18dca7821d4a3ec86421215d8fca857 (MD5) eap563_udesa_huhe_alb3_004.b.jpg: 1949806 bytes, checksum: 23d93ea6998ce6a08f8d89e4159c7b1 (MD5) eap563_udesa_huhe_alb3_001.a.tif: 29250522 bytes, checksum: d5f18baa8d307d6058fdaabe845c786d (MD5) eap563_udesa_huhe_alb3_001.a.tif.fits.xml: 8055 bytes, checksum: dc813f9e4870883fb25ca54b74fa5cca (MD5) eap563_udesa_huhe_alb3_001.b.tif: 30651520 bytes, checksum: 957f71fc57ae02ebad72d71d06467f75 (MD5) eap563_udesa_huhe_alb3_001.b.tif.fits.xml: 8055 bytes, checksum: 56aa92f24118ab7bed2a6a648897248 (MD5) eap563_udesa_huhe_alb3_002.a.tif: 29772074 bytes, checksum: c69afc85e072535a14ae27eadb4e3d24 (MD5) eap563_udesa_huhe_alb3_002.a.tif.fits.xml: 8055 bytes, checksum: 5d9cae2679f191e2a48f4d0aca1787e (MD5) eap563_udesa_huhe_alb3_002.b.tif: 29314722 bytes, checksum: a002ed5403d0b5ea682d8abdbb18980 (MD5) eap563_udesa_huhe_alb3_002.b.tif.fits.xml: 8055 bytes, checksum: 21320882cbb8a36b8c51a145fe314235 (MD5) eap563_udesa_huhe_alb3_003.a.tif: 30619688 bytes, checksum: b242df09ace52b95e1ad64081d8660e5 (MD5) eap563_udesa_huhe_alb3_003.a.tif.fits.xml: 8055 bytes, checksum: 3b22e0e75b6d6929a34c4bfc0e0ebc09 (MD5) eap563_udesa_huhe_alb3_003.b.tif: 28834834 bytes, checksum: ade0cfc48c775d70d8de1113c71a9d84 (MD5) eap563_udesa_huhe_alb3_003.b.tif.fits.xml: 8055 bytes, checksum: 60067230a9dc81c33e4eaa06d4d6e36 (MD5) eap563_udesa_huhe_alb3_004.a.tif: 29388896 bytes, checksum: 846d8549fa4e1da1f8ef4a9eb89867f2 (MD5) eap563_udesa_huhe_alb3_004.a.tif.fits.xml: 8055 bytes, checksum: 80c4f4de10f8b11cc810d51fd0e30ca0a (MD5) eap563_udesa_huhe_alb3_004.b.tif: 28379390 bytes, checksum: f26e851765144923ac74c4e39dd99ea5 (MD5) eap563_udesa_huhe_alb3_004.b.tif.fits.xml: 8055 bytes, checksum: 8bd0a9c180eb756f81c7d571a4fe655d (MD5) license.txt: 520 bytes, checksum: 0bb7393c55313db36325e242984ee5e (MD5)		en
dc.description.sponsorship Endangered Archives Program		es
dc.format.mimetype tiff		es
dc.relation.ispartof hume hermanos		es
dc.relation.ispartof albumes_fotografias		es
dc.relation.haspart eap563_udesa_huhe_alb3_001.a		es
dc.relation.haspart eap563_udesa_huhe_alb3_001.b		es
dc.relation.haspart eap563_udesa_huhe_alb3_002.a		es
dc.relation.haspart eap563_udesa_huhe_alb3_002.b		es
dc.relation.haspart eap563_udesa_huhe_alb3_003.a		es
dc.relation.haspart eap563_udesa_huhe_alb3_003.b		es
dc.relation.haspart eap563_udesa_huhe_alb3_004.a		es
dc.relation.haspart eap563_udesa_huhe_alb3_004.b		es
dc.subject albumes		es
dc.subject fotografias - trenes - retratos - vias detren		es
dc.title Album3		es
dc.type Images		es
dc.publisher.digital Universidad de San Andrés		es

Aparece en las colecciones: [5: Albumes de fotografías](#)

Ficheros en este ítem:

Fichero	Descripción	Tamaño	Formato
eap563_udesa_huhe_alb3_001.a.jpg		2.98 MB	JPEG Visualizar/Abbrir
eap563_udesa_huhe_alb3_001.b.jpg		2.28 MB	JPEG Visualizar/Abbrir
eap563_udesa_huhe_alb3_002.a.jpg		1.99 MB	JPEG Visualizar/Abbrir
eap563_udesa_huhe_alb3_002.b.jpg		1.94 MB	JPEG Visualizar/Abbrir
eap563_udesa_huhe_alb3_003.a.jpg		2.18 MB	JPEG Visualizar/Abbrir
eap563_udesa_huhe_alb3_003.b.jpg		1.96 MB	JPEG Visualizar/Abbrir
eap563_udesa_huhe_alb3_004.a.jpg		2 MB	JPEG Visualizar/Abbrir
eap563_udesa_huhe_alb3_004.b.jpg		1.9 MB	JPEG Visualizar/Abbrir

Mostrar el registro sencillo del ítem

View Statistics

Los ítems del repositorio están protegidos por copyright, con todos los derechos reservados, a menos que se indique lo contrario.

W3C

Proyecto liderado por la [Biblioteca Max von Buch](#) e implementado por el equipo de [Tecnología Informática](#) de la [Universidad de San Andrés](#) · [Comentarios](#) · [Seguinos en](#) [s](#) [f](#)

[http://repositorio.udesa.edu.ar/jspui/handle/10908/847?mode=full&submit_simple=Mostrar+el+registro+Dublin+Core+completo+del+%C3%ADItem+Thu+Oct+24+2013+16:16:35+GMT-0300+\(Hora+estándar+de+Argentina\)](http://repositorio.udesa.edu.ar/jspui/handle/10908/847?mode=full&submit_simple=Mostrar+el+registro+Dublin+Core+completo+del+%C3%ADItem+Thu+Oct+24+2013+16:16:35+GMT-0300+(Hora+estándar+de+Argentina))

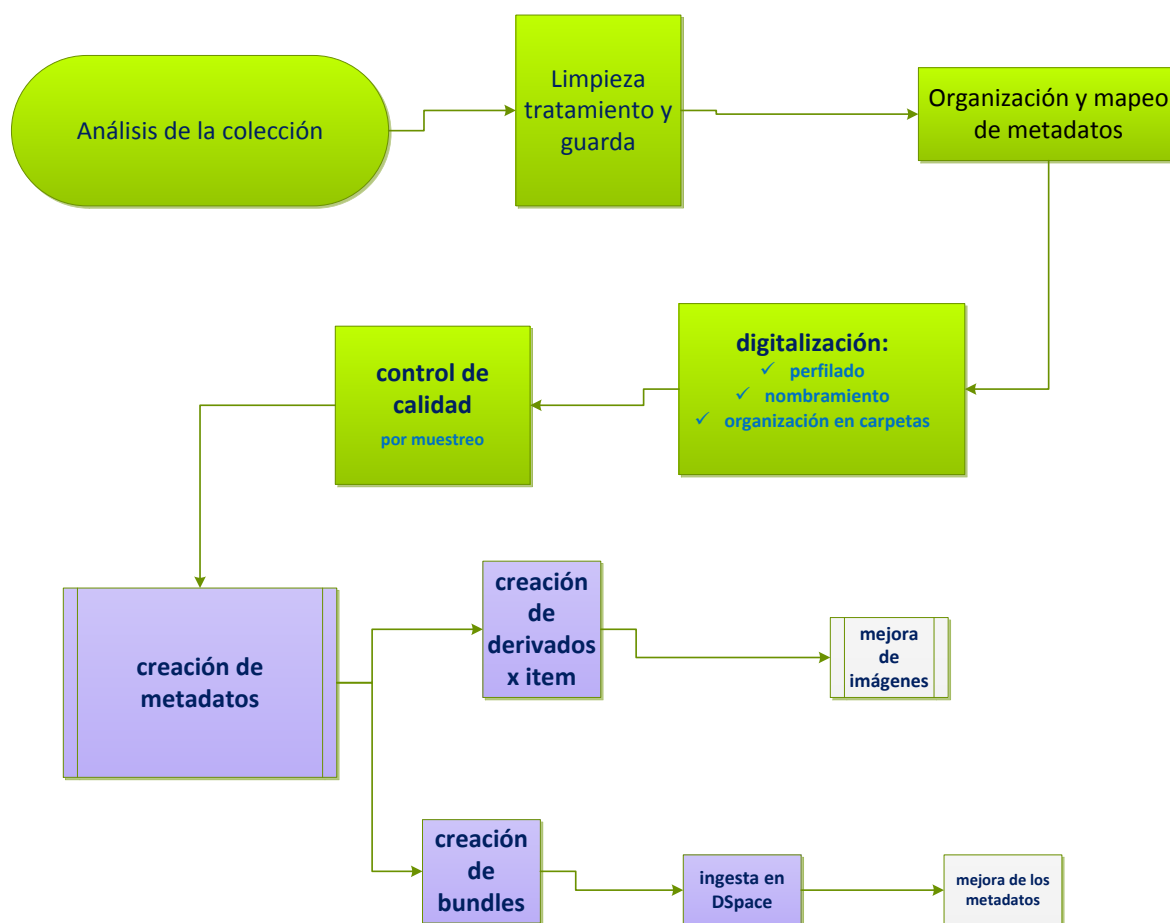
Figura 19: Archivo XML de un ítem en *DSpace*

Página 22 de 27

8- Flujo de trabajo

El flujo de trabajo es una representación gráfica de los diferentes pasos y etapas de todo el proceso. En él se indican las diversas tareas así como las relaciones de unas con otras y la forma en la que se encadenan para alcanzar el resultado final. Este flujo de trabajo, que se diseña durante una etapa de prueba y planificación, puede automatizarse una vez terminado. Esto justamente es lo que hace el programa que diseñó el informático, es decir que une los pasos desde la digitalización hasta que las imágenes y sus metadatos de preservación y descriptivos son almacenados en el repositorio.

Flujo de trabajo para preservación digital



9.- Medidas de preservación: réplica o back up

Una medida esencial de la preservación digital es la realización de réplicas o back-up periódicos de la colección de objetos digitales y sus metadatos. En la Universidad de San Andrés esto se realiza diariamente. Se trata de un backup completo, sobrescribiendo el anterior. Es decir que este back up no se hace sobre las novedades, los cambios o los ítems nuevos. El back-up se efectúa tanto

desde la base de datos como desde el *assetstore* (almacén de activos) de **DSpace** y se graba en una cinta magnética (LT03). Todo el proceso es automático, excepto el cambio (o rotación) de la cinta. Si no se cambia la cinta, se hace siempre sobre la misma. Al hacer el cambio de cinta, se guarda la que ha sido utilizada en una caja ignífuga y se pone en el servidor la siguiente cinta de la rotación. La Universidad de San Andrés cuenta con una rotación de 5 cintas, que son resguardadas en la caja ignífuga, la cual se encuentra en un ambiente distinto al del servidor. Lo ideal es que la caja ignífuga esté alojada en otro edificio, incluso, en otra localidad. En la Universidad, por el momento, se está planificando el traslado de la caja a otro edificio.

10.-Softwares y estándares utilizados en el proyecto

SOFTWARE/ESTÁNDAR	USO	DESCRIPCION	CODIGO ABIERTO
Epson Scan Ver.3.494	Digitalización	Se utiliza para la configuración de parámetros como resolución, tamaño Y enfoque en el momento de la digitalización.	NO
Spyder 4 PRO 4.5.4	Digitalización	Se utiliza para la calibración del monitor.	NO
FastStone Image Viewer 4.6	Digitalización	Es un software utilizado para la visualización de imágenes. Es compatible con todos los principales formatos gráficos y formatos de cámara digital.	SI
Adobe Photoshop Element 9	Digitalización	Para convertir archivos de un formato a otro de forma individual o en lote. Por ejemplo creación de .JPG. También se utiliza para modificar las imágenes.	NO
Adobe Bridge CS6	Digitalización	Este software contiene herramientas que permiten convertir, renombrar y organizar imágenes.	NO
FITS (File Information Tool Set)	Extracción de metadatos	Se utiliza para la identificación, validación y extracción de metadata técnica para varios formatos de archivo. Este software fue creado por la Biblioteca de la Universidad de Harvard como sistema para emplearlo en su servicio de depósito digital.	SI
DSpace	Preservación de datos	Software que provee herramientas para la administración de colecciones digitales, y comúnmente es usada como repositorio institucional para la	SI

		preservación de datos.	
Bulk Submission -- Enhanced Macro (DRMC) OhioLink	Generación metadatos	Herramienta para generación automática y carga masiva de metadatos en DSpace	NO
EAD (Encoded Archival Description)	Mapeo	Estándar XML utilizado para codificar instrumentos de descripción archivística desarrollado por la Library of Congress y la SAA (Asociación de Archivistas Americanos)	No
Dublin Core	Mapeo	Estándar XML utilizado para codificar metadatos, elaborado y auspiciado por la DCMI (Dublin Core Metadata Initiative).	NO

11.- Links relacionados con el proyecto institucional de digitalización

Colecciones Especiales y Archivos

<http://biblioteca.udesar.edu.ar/CEyA/Presentacion>

Endangered Archives Programme. British Library

<http://eap.bl.uk/>

Proyecto digitalización del Fondo Hume Hermanos

http://eap.bl.uk/database/overview_project.a4d?projID=EAP563;r=20798

Curriculum Vitae Dra. María Esteva

<http://www.tacc.utexas.edu/staff/maria-esteva>

Programa Raíces. Ministerio de Ciencia, Tecnología e Innovación Productiva

http://www.mincyt.gob.ar/programas/index.php?Id_programa=8

12.-Bibliografía y links relacionados con el proceso de digitalización y preservación

Digitalización

<http://www.hugorodriguez.com/>

DSpace

<http://www.dspace.org/>

Creative Commons

<http://creativecommons.org>

ICA AtoM

<https://www.ica-atom.org/>

DUBLIN CORE

<http://dublincore.org/>

EAD

<http://www.loc.gov/ead/>

EXPORTACIÓN DE METADATOS

<https://sites.google.com/a/ohiolink.edu/drmc/bulk-submission/bulk-submission----alternate-e>

FITS

<https://code.google.com/p/fits/>

CREATIVE COMMONS

<http://creativecommons.org/>

BRASILIANA

<http://www.brasiliana.usp.br/bbd/handle/1918/00624510>

LIGHTBOX

<http://explorer.our.dmu.ac.uk/2011/07/28/lightboxes-thumbnails-and-slideshows-in-dspace-kultur-the-challenge/>

TUTORIALES

Digitization Guidelines for Photographs and Textual Documents

Moving theory into Practice. Digital Imaging Tutorial

<http://www.library.cornell.edu/preservation/tutorialcontents.html>

LIBROS

Moving Theory into Practice. Digital Imaging for Libraries and Archives. Anne R. Kenney – Oya Y.Rieger. 2000 by Cornell University Library

Contactos:

Silvana Lucía Piga

spiga@udesa.edu.ar

Natalia Westberg

nwestberg@udesa.edu.ar

Para ambos contactos

archivo@udesa.edu.ar

Teléfono: 011 4725 7043